

DISTRIBUTED AUTHORIZING AND VERSIONING

Gail Kaiser • Columbia University • kaiser@cs.columbia.edu
 Jim Whitehead • University of California at Irvine • ejw@ics.uci.edu

The World Wide Web has made retrieving multimedia information on the Internet extremely easy, accessible now to millions. Authoring and publishing Web content, however, remain a difficult—some would say, onerous—task.

At present, authors have three options, the most prevalent of which is to write to the storage medium of the server. This typically involves using operating system services to write to a filesystem. This implies having some form of local access, often a login session on a machine connected via a local area network to a network file system containing the content exported by the Web server. A variation on this approach is to use FTP to remotely write content to a server's filesystem. The third option is to use HTTP* capabilities to write directly to the namespace of the Web server.

CURRENT PROBLEMS

Authors using filesystem facilities to publish their Web content must contend with several tedious issues:

- They must understand the mapping between the names used by the storage medium of a server (typically the files and directories of a filesystem) and the URLs that the server understands.
- They must manually perform resource locking to ensure the content isn't overwritten by another author working simultaneously.
- They must have login privileges on the system hosting their content in order to list directory contents or create new directories to subdivide the name space.
- They have no way of recording information about the content (metadata), such as author, maintaining organization, and so on, for non-HTML content.

Using the facilities of HTTP to perform remote authoring removes the need to understand the mapping between the filesystem and the URL namespace, and hence improves on working in the filesystem. While HTTP-based remote authoring may require a password for network security, it is typically easier to create a new user on an HTTP server than it is to create a new account on a filesystem. Unfortunately, working with HTTP does not solve the problems with resource locking, nor does it provide capabilities for recording metadata on resources. Additionally, working exclusively with HTTP facilities engenders some new concerns.

- Authors must work only with the most recent revision of a document due to the Web's lack of standard versioning support.
- If the content is processed by the server before being transmitted back to a client, login privileges are required to access the unprocessed content source. Many Web servers support Server Side Include functionality that can be used, for example, to automatically insert the current date into a Web page.

These problems have thus far prevented the Web from being widely used as a platform for distributed authoring of Web content. At present there are many point solutions attempting to address them. While America Online AOLpress,* Microsoft Frontpage,* Netscape Navigator Gold,* and W3C Amaya* all offer pushbutton publishing directly to a URL using HTTP PUT, they offer varying, often proprietary means for preventing the "lost update problem." AOLpress also offers users the ability to list the contents of a directory, and create new directories (levels in the URL hierarchy), using nonstandard extensions to HTTP. Mortice Kern Systems' Web Integrity* product provides locking and versioning of Web content by using a nonstandard URL decoration scheme to encode versioning commands.

THE FORMATION OF WEBDAV

The need for interoperability of distributed Web content authoring tools led to the formation in 1996 of the working group on Distributed Authoring and Versioning on the World Wide Web.* WEBDAV has recently become a working group within the Internet Engineering Task Force* and also works cooperatively with the World Wide Web Consortium.* Active participants include several authors of the HTTP specification, organizations developing core Web technologies, document management companies, configuration management vendors, HTML authoring tool designers, as well as others interested members of the Web community. (You can join the mailing list for the group by sending e-mail with subject "subscribe" to w3c-dist-auth-request@w3org.*)

NEW EXTENSIONS TO HTTP 1.1

To address the existing deficiencies of the Web for distributed authoring, the group decided to define the following extensions to the HTTP 1.1 protocol.

Locking. An exclusive write lock capability prevents two (or more) users from overwriting each other's work. A lock discovery mechanism allows authors to find out if

any locks exist on a Web resource. Partial resource locking is also supported, allowing byte ranges within a document to be locked. Partial locking can be used to support collaborative editing of word-processing documents that have chunked file formats, where each chunk can be separately locked. Since the Web is designed so that no lock is required to read a Web page, there is no concept of a read lock. This does raise the possibility that the contents of a resource may change without warning if a write lock is not owned on the resource.

Metadata. Typed links may be defined on Web resources, creating simple attribute-value pairs consisting of the link type and the contents of the destination of the link. This allows arbitrary metadata to be associated with a Web resource.

Name space management. Operations to move and copy Web resources, and to create and list new container objects (directories), allow users to remotely manage the name space of their Web resources.

Access to unprocessed source. A predefined link on a resource points to a location where a client may retrieve the source of a Web resource without processing by the server.

Versioning. Simple check-out and check-in operations, along with the ability to retrieve the version history and older versions of a resource, and to attach a comment on check-out or check-in, are supported. Other versioning capabilities, such as having multiple branches, are supported by the protocol but may not be supported by the underlying versioning system.

DISTRIBUTED AUTHORING'S FUTURE

These capabilities are obviously useful for remote editing of HTML documents. They also enable distributed editing of mainstream application media types such as word-processor files and spreadsheets, since HTTP operations are independent of the Web resource's media type. Furthermore, this functionality provides a simple interface for accessing not only a Web server's resources, but also the files in a document or configuration management system. Distributed authoring and versioning also facilitate the formation of virtual organizations. On the Internet, a virtual organization is typically formed by a group of people interested in working on a particular project or goal who form a mailing list and a Web site to promote the group and to serve as a collection point for related information. An example is the Apache* group of widely dispersed developers who created the market-leader Apache Web server. Apache has an internal developers' mailing list, as well as a public Web site where their software can be downloaded. Unfortunately, today's virtual organizations have few resources for performing collaborative work over the Internet using standard tools.

Distributed authoring and versioning may provide a standard collaboration framework, supported by interoperable

products from many vendors. By setting up a Web server capable of distributed authoring and versioning, a virtual organization can use the server like a simple network filesystem. Members from around the globe can use their local applications to edit documents on the remote server. More typical organizations can also benefit from distributed authoring and versioning, for example, by allowing an out-of-office sales or support force to work collaboratively on documents stored at the home office.

Distributed authoring and versioning is moving the Web away from being simply a medium for publishing content toward a new role as a globally distributed read/write medium. The Web is thus becoming a simple, globally distributed, versionable filesystem, ready for a wide range of collaborative uses.

DISTRIBUTED COLUMNING NOW

Halfway through writing this column, we decided to try using distributed authoring technology to complete it. We put the column on an AOLserver, then edited it in place using AOLpress. One author was in New York City, the other in Irvine, California, with the server located in Santa Barbara, California, all three belonging to different organizations. We hoped to use the proprietary locking facilities of AOLpress to prevent overwrites. For one author, the experience was similar to using a word processor: The author opened the column by giving a URL using the File/Open menu, locked the document using the Tools/Administer Page/Lock menu, performed the edits, saved the final result using File/Save, then unlocked the document using Tools/Administer Page/Unlock. The other author was initially able to edit the column in place, but was later unable to lock the column and had to bypass the system by loading and saving a temporary copy of the page using another web browser. Deadline pressure did not allow us to work through this problem with the extremely responsive staff at America Online.

Though disappointing, this editing scenario does show the potential value of distributed authoring. However, it does not demonstrate interoperable distributed authoring since both authors were using an America Online client and server—in fact, the authors initially attempted to use HTML editors produced by different vendors, and experienced difficulty with cross-vendor publishing to the AOLserver. Future tools based on the WEBDAV specification will not have these interoperability problems.

—If you would like to request coverage in this column of a particular subarea of collaborative work, or better yet guest coauthor a column on a specific topic, contact Gail Kaiser.

*URLs from these pages

HTTP • www.ics.uci.edu/pub/ietf/http
 AOL Press • www.aolpress.com/press/index.html
 Frontpage • www.microsoft.com/frontpage
 Navigator Gold • home.netscape.com/comprod/products/navigator/gold
 Amaya • www.w3.org/pub/WWW/Amaya
 Web Integrity • www.mks.com/solution/ie/
 WEBDAV • www.ics.uci.edu/~ejw/authoring/